Comput Geosci (2017) 21:645–663
DOI 10.1007/s10596-017-9641-4

ORIGINAL PAPER

Flow-based dissimilarity measures for reservoir models: a spatial-temporal tensor approach

Edwin Insuasty[1] · Paul M. J. Van den Hof[1] · Siep Weiland[1] · Jan-Dirk Jansen[2]

Received: 22 August 2016 / Accepted: 16 March 2017 / Published online: 8 April 2017

Abstract In reservoir engineering, it is attractive to characterize the difference between reservoir models in metrics that relate to the economic performance of the reservoir as well as to the underlying geological structure. In this paper, we develop a dissimilarity measure that is based on reservoir flow patterns under a particular operational strategy. To this end, a spatial-temporal tensor representation of the reservoir flow patterns is used, while retaining the spatial structure of the flow variables. This allows reduced-order tensor representations of the dominating patterns and simple computation of a flow-induced dissimilarity measure between models. The developed tensor techniques are applied to cluster model realizations in an ensemble, based on similarity of flow characteristics.

Keywords Reduced-order modeling · Tensor decompositions · Tensor algebra · Flow characterization

✉ Paul M. J. Van den Hof
p.m.j.vandenhof@tue.nl

Edwin Insuasty
e.g.insuasty.moreno@tue.nl

Siep Weiland
s.weiland@tue.nl

Jan-Dirk Jansen
j.d.jansen@tudelft.nl

[1] Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, Den Dolech 2, 5612AZ, Eindhoven, The Netherlands

[2] Department of Geoscience and Engineering, Delft University of Technology, Stevinweg 1, 2628, CN Delft, The Netherlands

1 Introduction

The increasing demand for energy has encouraged the use of improved production strategies for conventional oil and gas resources. In this context, several studies have indicated a significant scope for reservoir model-based life-cycle optimization of ultimate recovery or net present value (NPV), especially when combined with computer-assisted history matching leading to a closed-loop reservoir management (CLRM) approach (see, e.g., [18, 29] or [10]). In this CLRM approach, the assessment of the value of information of different resources becomes an important feature (see, e.g., [6]). To properly account for the effect of geological uncertainty, it is important to perform both the history matching, life-cycle optimization and value of information assessment on the basis of several realizations of the reservoir model. The combination of iterative (large-scale) optimization, and history matching, with the need to use multiple model realizations makes CLRM into a computationally very demanding process for realistically-sized reservoir models. Particularly, for the CLRM framework, one would like to discriminate between realizations which are representatives of the different types of flow responses. For this reason, oil companies have used very few realizations which are often selected manually, to achieve robustness in their operational strategies (see, e.g., [28]). The selection of representative models has become a relevant issue for the practice of reservoir engineering. In other words, there is a need for a dissimilarity measure between reservoir realizations that is relevant for model-based operation of oil reservoirs.

There are several options for discriminating between model realizations, on the basis of either static or dynamic properties of the reservoir models. [35] and [8] have used the permeability fields as a measure of dissimilarity.

This has been done by defining a metric space to compare and cluster geological models that share common geological features. These dissimilarity measures based on *static* permeability or porosity properties are known to be quite different from measures applied to the *dynamic* behavior of the reservoir models, reflected in the corresponding flow patterns, as, e.g., the evolution of oil saturation over the life cycle of the reservoir. To exemplify this, upscaling of high-resolution geological models, as is presented in [13], shows that dissimilar reservoir models may have similar dynamical performance in terms of flow dynamics. At the same time, reservoir models that are close in geological properties can have essentially different flow patterns, and therefore different behavior from a dynamic operation point of view.

From a production optimization perspective, the use of NPV, generated at the end of a certain production period, could be a natural basis for a control-relevant dissimilarity measure between model realizations. If we would restrict attention to life cycle production optimization under predefined production strategy and well configurations, this might be true. However, when reservoir models are to be used also for testing new production strategies, as well as well-placement, infill-drilling and re-completion plans, the NPV measure is considered to be too coarse to distinguish between essential dynamic properties of the models. It is well known that the NPV is not able to capture the relevant aspects of the reservoir flow patterns associated with a particular production strategy, in other words: two essentially different geological models could lead to the same NPV under similar production strategies, but on the basis of essentially different flow patterns.

In [30] and in [31], the total oil production and water rates are used as dissimilarity measures to assess the dynamical responses of different reservoir realizations. Although these measures are less coarse than the NPV, oil and water production rates (combined with pressure measurements) only provide local information of the flow behavior around the wellbore, which cannot be extrapolated to characterize the flow performance of spatial locations far from the production sites. Therefore, it suffers from similar limitations as measures based on NPV.

While streamline simulators have been used to generate a fast characterization of the cumulative production rates (see, e.g., [27, 32], and [30]), they have also led to a technique called *dynamic fingerprinting* ([41]), where streamline information (time-of-flight (TOF) or drainage time) is used to generate flow patterns that are, like a fingerprint, unique to each realization. Then, fingerprints are used to screen and cluster reservoir realizations with similar dynamical performance. The resulting dissimilarity measures are attractive for flow characterization, though they are merely simplified descriptors of the much more complex spatial-temporal reservoir flow patterns in terms of evolution of the flow variables (phase saturations, pressures, etc.) for a particular production strategy.

In this paper, we address the question whether we can use the full reservoir flow patterns as the dissimilarity measure between reservoir realizations. Reservoir flow patterns are numerical solutions of the pressure and transport partial differential equations (PDEs) ([4]), and they represent the temporal evolution of the dependent variables in the spatial domain (typically $10^5 \sim 10^6$ grid blocks) of the reservoir. Therefore, the discrete-time trajectories for pressure, saturation, temperature, etc., are usually large-scale data structures, with dimensions induced by the number of grid cells of the reservoir model. The large dimensionality of these structures would make them unsuitable to serve as a dissimilarity measure for performing model discrimination and clustering, and therefore reduced-order representations are necessary. Previously, [9, 26] and [23] have used the singular value decomposition (SVD) and POD model order reduction techniques to arrive at low dimensional representations of flow variables, while [41] have used SVD to represent the fingerprints through a reduced set of basis functions. However, the SVD approach has some limitations. As the reservoir flow patterns are stacked in vectors, the natural spatial-temporal structure of the reservoir is lost. This may have serious implications when characterizing flow profiles in low-dimensional spaces, as some information related to the spatial correlations is lost during the vectorization scheme, see [16].

In this paper, we develop a tensor approach for efficient storage of reservoir flow patterns in a multidimensional array. This creates a clear separation of the spatial, temporal, and flow variables coordinates, and allows for reduced-order representations using basis functions in each of the separate coordinates, thereby appropriately maintaining spatial correlation structures. With an additional extension of the tensor coordinates, it will even allow for describing the flow characteristics of an *ensemble* of models. Tensor decompositions and tensor analysis constitute a largely unexplored subject in reservoir engineering. For the characterization of geological parameters, [1] and [2] have performed low rank approximations of permeability fields using a tensor decomposition, and [14] have used these representations for efficient history matching. For the reduction of dynamical complexity of reservoir models, [16] have utilized this framework for constructing reduced-order dynamic models. In our current paper, we apply state-of-the-art tensor decomposition techniques to characterize flow profiles in low-dimensional spaces. The corresponding reduced-order representations will be analyzed for their suitability to calculate distance measures between models, and for subsequent distance visualization and model clustering.

The paper is organized as follows: In Section 2, we introduce the notions of flow-based dissimilarity measures and

the state-of-the-art technology for flow characterization. In Section 3, we present the benefits of exploiting the spatial structure and correlations of the reservoirs by using a tensor formulation and we introduce our spatial-temporal approach for the flow characterization through dissimilarity measures. In Section 4, we present a tensor-based workflow for the flow characterization of an ensemble of reservoir models. In Section 5, we evaluate the performance of the workflow for flow characterization using flow-based dissimilarity measures.

2 Flow-based dissimilarity measures

2.1 Introduction

In water-flooding, the temporal evolution of the oil saturation (and in particular the oil-water front) provides sensible information for well placement and for the design of schedules for the well controls in order to optimize production. Hence, the reservoir flow patterns are the variables with physical interpretation that best describe the dynamic properties of the hydrocarbon reservoir, and we can conceptually state that two reservoir realizations are similar with respect to their dynamical performance if for a particular operational strategy, the generated reservoir flow profiles are similar.

The variable $s(x, t, u)$ will be used in this paper to represent the flow-related variable, with a spatial coordinate $x \in \mathbb{R}^2$, time $t \in \mathbb{R}$, and operational strategy u. In most cases, s will correspond to the oil saturation in each (spatial) grid block, although other variables (e.g., pressure, time-of-flight, drainage time) could be included too. They are the solutions of the underlying model's multiphase flow equations through their corresponding PDE's, and the result of a particularly chosen operational strategy of water injection and control valve settings, reflected by the variable u, [17]. In this section, we elaborate on the concept of model distances based on reservoir flow patterns.

2.2 Dissimilarity measures and distance functions

When quantifying flow-based dissimilarities between reservoir models, one should consider the use of distance functions. A distance function defines the separation between two elements in a set (the set of reservoir flow responses) and it induces a metric space, where the distance between two different reservoir models is an indicator of their dissimilarity in the dynamical response. There are many functions to compute the distance between two objects: the Euclidean distance, standardized Euclidean, Chebyshev distances, and many more. [34] have used the Hausdorff distance to measure the dissimilarity of geometry for reservoir realizations, and [27] have used connectivity distances based on streamlines. If s_1 and s_2 are the flow-related variables corresponding to two different models, a natural dissimilarity measure to consider is a quadratic distance measure:

$$d(s_1, s_2) = \sqrt{\sum_{k=1}^{K}\sum_{i=1}^{I}\sum_{j=1}^{J} \|s_1(x_{ij}, t_k, u) - s_2(x_{ij}, t_k, u)\|^2} \tag{1}$$

where K is the total number of time steps; I, J are the number of grid cells in each spatial dimension, and the two models are operated with the same operational strategy $u(t_k)$. For brevity of notation, we will often discard the dependency of $s(x, t_k, u)$ on u and simplify the notation to $s(x, t_k)$ whenever there is no risk of confusion. The underlying spatial domain is assumed to be rectangular with Cartesian grid. The temporal evolution of the flow variables $s(x, t_k)$ over all grid cells is a collection of high-dimensional state variables and generally requires the use of huge computational resources for storage, function evaluations, the evaluation of distances as in (1) and its subsequent use for visualization and model clustering. In the next section, a method for the low-dimensional representation of $s(x, t_k)$ is described.

2.3 Low dimensional representations and flow-based distances through SVD

Compact representations of $s(x, t_k)$ are important for an efficient and fast numerical calculation of distance measures, see [41]. A typical way to construct lower dimensional representations of $s(x, t_k)$ is obtained by utilizing a basis function expansion for the set of flow variables over all grid cells:

$$s(x, t_k) = \sum_{i=1}^{\hat{R}} \sigma_i(t_k)\varphi_i(x), \tag{2}$$

where the basis functions $\varphi_i(x)$, for $i = 1, \cdots, \hat{R}$ can be selected to be the most informative spatial patterns in the flow response. If $\hat{R} \ll N$, where $N = I \cdot J$ is the number of grid cells, we say that the reservoir flow pattern $s(x, t_k)$ is characterized in a low-dimensional space by the coefficients $\sigma_i(t_k)$ and by the basis functions $\varphi_i(x)$, for $i = 1, \cdots, \hat{R}$. The classical technique for obtaining this representation is through principle component analysis (PCA) and the use of singular value decompositions (SVD), [15]. To this end, the dynamic variables in the grid are represented as $(I \cdot J) \times 1$ vectors, denoted as $\mathbf{x}_k$, with elements $s(x_{ij}, t_k)$ at a particular time moment t_k. A number K of these snapshot vectors $\mathbf{x}_k$ is collected at time instants $t_1, \cdots, t_K$, where K may be

less than or equal to the total number of simulation time steps. With $N = I \cdot J$, this results in a $N \times K$ matrix of data points

$$\mathbf{X} = \left[\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_K \right], \tag{3}$$

which is decomposed using SVD through:

$$\mathbf{X} = \Phi \Sigma \Psi^\top = \sum_{r=1}^{R} \sigma_r \varphi_r \psi_r^\top = \sum_{r=1}^{R} \sigma_r \varphi_r \otimes \psi_r, \tag{4}$$

where Φ and Ψ are $N \times N$ and $K \times K$ orthogonal matrices containing the left and right singular (column) vectors φ_r and ψ_r, Σ is an $N \times K$ rectangular diagonal matrix that has the ordered singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_R \geq 0$ on its main diagonal, R is the rank of $\mathbf{X}$, and $\otimes$ denotes the tensor or outer product over a vector space. Usually, $R \leq K \ll N$ in a typical reservoir simulation application. The last equality in (4) indicates that $\mathbf{X}$ can be decomposed as the sum of R rank-one matrices $\Theta_r = \varphi_r \otimes \psi_r$. In particular, every individual snapshot vector $\mathbf{x}_k$ can be written as:

$$\mathbf{x}_k = \sum_{r=1}^{R} \sigma_r \varphi_r \psi_r^\top \mathbf{e_k} = \sum_{r=1}^{R} \alpha_r^k \varphi_r, \tag{5}$$

where $\mathbf{e_k}$ is the kth standard unit vector in $\mathbb{R}^K$ and $\alpha_r^k := \sigma_r \psi_r^\top \mathbf{e_k}$ is a real-valued coefficient. The R left singular vectors φ_r, $r = 1, \ldots, \hat{R}$ then characterize the spatial correlations of the original snapshot matrix $\mathbf{X}$, ordered in decreasing relevance, and allows a low-dimensional approximation $\hat{\mathbf{x}}_k = \sum_{r=1}^{\hat{R}} \alpha_r^k \varphi_r$, with $\hat{R} < R$, of the reservoir flow patterns and fingerprints in terms of the coefficients α_r^k. Using (5), the Euclidean distance between the ith and jth snapshots $\mathbf{x}_i, \mathbf{x}_j$ is:

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| = \sqrt{\left(\sum_{r=1}^{R} (\alpha_r^i - \alpha_r^j) \varphi_r \right)^2} = \sqrt{\sum_{r=1}^{R} (\alpha_r^i - \alpha_r^j)^2}. \tag{6}$$

Now, let us consider the low-rank approximation $\widehat{\mathbf{X}} = \left[\hat{\mathbf{x}}_1 \ \hat{\mathbf{x}}_2 \ \cdots \ \hat{\mathbf{x}}_K \right]$ of $\mathbf{X}$, which can be obtained by decomposing (4) as

$$\mathbf{X} = \widehat{\mathbf{X}} + \overline{\mathbf{X}} = \underbrace{\sum_{r=1}^{\hat{R}} \sigma_r \varphi_r \otimes \psi_r}_{\widehat{\mathbf{X}}} + \underbrace{\sum_{r=\hat{R}+1}^{R} \sigma_r \varphi_r \otimes \psi_r}_{\overline{\mathbf{X}}}, \tag{7}$$

where $\hat{R} < R$ is the approximation order and where $\overline{\mathbf{X}} = \mathbf{X} - \widehat{\mathbf{X}}$ is the approximation error. For any $\hat{R} < R$, the Frobenius norm of the error

$$\|\mathbf{X} - \widehat{\mathbf{X}}\|_F = \sqrt{\sum_{r>\hat{R}} \sigma_r^2}$$

is minimal over all rank $\hat{R}$ approximations of $\mathbf{X}$. The approximate representations $\{\hat{\mathbf{x}}_k\}_{k=1,\cdots K}$ of the reservoir flow pattern can now be used as a basis for measuring dissimilarities between models, where the appropriate calculations can be performed on the basis of the coefficients α_r^k for $r = 1, \cdots \hat{R}$ and $k = 1, \cdots K$. Let us consider the low-dimensional characterization of the flow patterns in terms of the coefficients α_r^i, α_r^j for $r = 1, \ldots, \hat{R}$, then the approximated dissimilarity is:

$$\widehat{d}_{ij} = \sqrt{\sum_{r=1}^{\hat{R}} (\alpha_r^i - \alpha_r^j)^2}, \tag{8}$$

where $\widehat{d}_{ij}$ is the (i, j)th element of a matrix $\widehat{\mathbf{D}} \in \mathbb{R}^{K \times K}$ of all approximate distances.

2.4 Discussion

The SVD-based approach presented in Section 2 has been adopted in industrial practice, [41], but it has some limitations. Through the vectorized form in which flow variables are stored, the spatial-temporal structure of the reservoir is lost. This may have serious implications when characterizing flow profiles in low-dimensional spaces. When SVD is applied to a snapshot matrix $\mathbf{X}$, the sets of orthonormal basis vectors $\{\varphi_r\}_{r=1}^{\hat{R}}$, $\{\psi_r\}_{r=1}^{\hat{R}}$, average the energy of solutions in time, and by definition, do not discriminate among spatial coordinates. This temporal averaging of the energy causes a loss of information for some of the relevant features of spatial coordination in the data, see [16]. For linear systems like single-phase flow problems, the spatial correlations are invariant in time and can be characterized analytically using concepts from system theory such as controllability and observability, see [36]. However, the nonlinearities induced by the multi-phase character of the problems may define time-variant correlations of the states, and the correlation between time and specific spatial direction is ignored by vectorizing the flow variables, in which case it can be attractive to clearly separate all spatial and temporal coordinates to maintain their own independent role when constructing approximations. In the next section, a methodology that overcomes the limitations of SVD methods for flow characterization is presented.

3 Spatial-temporal tensor methods for flow-based dissimilarity measures

3.1 Introduction

In this section, we develop a multidimensional approach to understand the dynamical similarities between reservoir models, which is based on tensor representations and decompositions of flow related variables. In addition, we incorporate this approach into a workflow for clustering of models with similar dynamical performance.

3.2 Tensor representations of reservoir flow patterns

In the previous section, we have constructed vectorized representations of the flow variables ($I \cdot J \times 1$ snapshot vectors). Alternatively, if the spatial grid has a Cartesian structure, one can collect K snapshot matrices $\mathbf{X}_k$ of size $I \times J$, and represent this data object in a three-dimensional array $\mathscr{S}$ of size $I \times J \times K$. That is, the reservoir flow data is represented as a multi-array $\mathscr{S} \in \mathbb{R}^{I \times J \times K}$. Such a multidimensional array is called a *tensor* and can be viewed as the natural generalization of vectors and matrices to higher dimensional objects. For a 2D saturation field that evolves over time, a three-dimensional array is schematically depicted in Fig. 1.

A key advantage of multidimensional data objects is that they keep the spatial structure of the Cartesian grid intact. A disadvantage of the use of tensors is that their algebraic properties are more complicated and that numerical tools for tensor operations are less developed. In general, tensors are multilinear generalizations of algebraic objects such as vectors and matrices, and there exist suitable extensions of concepts such as decompositions, basis functions and spectral expansions to the multilinear case. In the next subsection, we describe the basic concept of tensor decompositions as an extension to the concept of matrix decomposition.

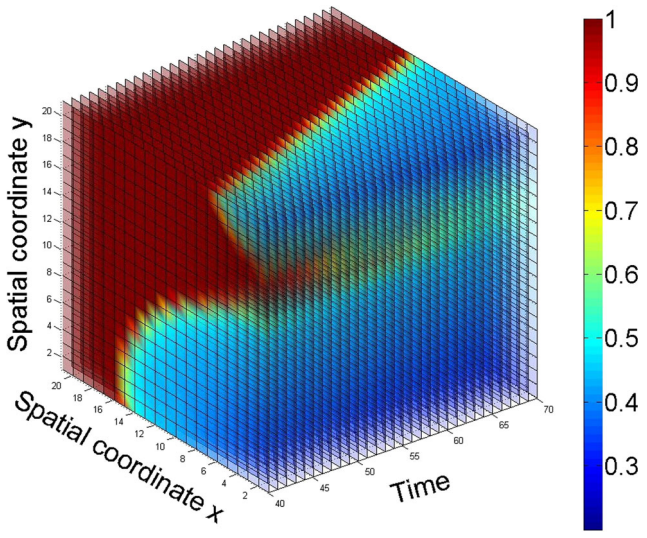

Fig. 1 Schematic of the tensor representation of a reservoir flow pattern. Axes represent spatial-temporal coordinates. Color-scale corresponds to oil saturation

3.3 Tensor decompositions and approximation

In analogy to the matrix decomposition in Eq. 4, the tensor $\mathscr{S}$ can be decomposed as a Tucker type decomposition ([22]):

$$\mathscr{S} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \sigma_{ijk}\, \varphi_i \otimes \psi_j \otimes \chi_k = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \sigma_{ijk}\Theta_{ijk}, \tag{9}$$

where the scalars $\sigma_{ijk} \in \mathbb{R}$ are the elements of the so called *core tensor* of size $I \times J \times K$ and where

$$\Theta_{ijk} := \varphi_i \otimes \psi_j \otimes \chi_k$$

is the outer product of vectors $\varphi_i \in \mathbb{R}^I$, $\psi_j \in \mathbb{R}^J$ and $\chi_k \in \mathbb{R}^K$. This makes Θ_{ijk} a rank-one three-way tensor. In any such representation, the sets $\{\varphi_i\}_{i=1}^I$, $\{\psi_j\}_{j=1}^J$ and $\{\chi_k\}_{k=1}^K$ are usually taken as a *basis* of $\mathbb{R}^I$, $\mathbb{R}^J$, and $\mathbb{R}^K$, respectively, and the Tucker decomposition (9) is viewed as a representation of the tensor with respect to these bases. Similar to the matrix case, tensors can be interpreted as multilinear mappings. More precisely, the rank-one three-way tensor Θ_{ijk} is a mapping $\Theta_{ijk} : \mathbb{R}^I \times \mathbb{R}^J \times \mathbb{R}^K \to \mathbb{R}$ defined as

$$\Theta_{ijk}(\varphi, \psi, \chi) := \langle \varphi_i, \varphi \rangle \langle \psi_j, \psi \rangle \langle \chi_k, \chi \rangle$$

which is a product of inner products in $\mathbb{R}^I$, $\mathbb{R}^J$ and $\mathbb{R}^K$. At this stage, it is important to observe that the mapping Θ_{ijk}, defined in this way, is linear in each of its argument. Moreover, if the bases $\{\varphi_i\}_{i=1}^I$, $\{\psi_j\}_{j=1}^J$ and $\{\chi_k\}_{k=1}^K$ are all orthonormal sets (that is, $\langle \varphi_{i'}, \varphi_{i''} \rangle = 1$ if $i' = i''$ and is zero otherwise for vectors $\{\varphi_i\}_{i=1}^I$), it follows that

$$\begin{aligned} \mathscr{S}(\varphi_{i_0}, \psi_{j_0}, \chi_{k_0}) &= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \sigma_{ijk}\Theta_{ijk}(\varphi_{i_0}, \psi_{j_0}, \chi_{k_0}) \\ &= \sigma_{i_0 j_0 k_0} \end{aligned} \tag{10}$$

for any triple of indices (i_0, j_0, k_0) with $1 \le i_0 \le I$, $1 \le j_0 \le J$ and $1 \le k_0 \le K$. In words, this says that the entries of the core tensor represent the tensor $\mathscr{S}$ when evaluated at its (orthonormal) basis vectors. If the bases are orthonormal, then this observation naturally identifies the entries $\sigma_{i_0 j_0 k_0}$ of the core tensor with the evaluation of the tensor $\mathscr{S}$ at its (i_0, j_0, k_0)th basis element. If the bases are non-orthonormal

bases, then the multilinear functional $\mathscr{S} : \mathbb{R}^I \times \mathbb{R}^J \times \mathbb{R}^K \to \mathbb{R}$ defined in (10) changes its representation. A graphical illustration of the tensor decomposition (9) is depicted in Fig. 2.

A low-rank approximation of $\mathscr{S}$ can be obtained by decomposing (9) according to $\mathscr{S} = \widehat{\mathscr{S}} + \overline{\mathscr{S}}$ where, for $\widehat{I} \le I, \widehat{J} \le J, \widehat{K} \le K$, the tensor

$$\widehat{\mathscr{S}} := \sum_{i=1}^{\widehat{I}} \sum_{j=1}^{\widehat{J}} \sum_{k=1}^{\widehat{K}} \sigma_{ijk} \Theta_{ijk} \tag{11}$$

is viewed as the approximation of $\mathscr{S}$ to its *modal-rank* $(\widehat{I}, \widehat{J}, \widehat{K})$ truncation and where $\overline{\mathscr{S}} := \mathscr{S} - \widehat{\mathscr{S}}$ is viewed as the corresponding approximation error. The size of the approximation error is measured in Frobenius norm and satisfies

$$\| \mathscr{S} - \widehat{\mathscr{S}} \|_F^2 \quad = \quad \| \mathscr{S} \|_F^2 - \sum_{i=1}^{\widehat{I}} \sum_{j=1}^{\widehat{J}} \sum_{k=1}^{\widehat{K}} \sigma_{ijk}^2 \tag{12}$$

provided that the bases $\{\varphi_i\}_{i=1}^I$, $\{\psi_j\}_{j=1}^J$ and $\{\chi_k\}_{k=1}^K$ are orthonormal sets.

Suppose that the above tensor decomposition is applied to the data corresponding to a two-dimensional rectangular saturation field that evolves over time. The mth sample $\mathbf{X}_m$ is then represented as a *matrix* of dimension $I \times J$. A number of K samples $\mathbf{X}_m$ is stored in an order-3 tensor $\mathscr{S}$ of size $I \times J \times K$. This tensor is approximated as in (11), and results in the approximate sample $\widehat{\mathbf{X}}_m$ of the saturation field defined by the order-2 tensor

$$\begin{aligned} \widehat{\mathbf{X}}_m &= \widehat{\mathscr{S}}(\cdot, \cdot, \mathbf{e_m}) = \sum_{i=1}^{\widehat{I}} \sum_{j=1}^{\widehat{J}} \sum_{k=1}^{\widehat{K}} \sigma_{ijk} \langle \chi_k, \mathbf{e_m} \rangle \, \varphi_i \otimes \psi_j \\ &= \sum_{i=1}^{\widehat{I}} \sum_{j=1}^{\widehat{J}} \alpha_{ij}^m \varphi_i \otimes \psi_j, \end{aligned} \tag{13}$$

where $\mathbf{e_m}$ is the mth standard unit vector in $\mathbb{R}^K$ and where $\alpha_{ij}^m := \sum_{k=1}^{\widehat{K}} \sigma_{ijk} \langle \chi_k, \mathbf{e_m} \rangle$ are real-valued coefficients in the expansion (13) of (rank 1) two-dimensional fingerprints of the saturation field. The coefficient α_{ij}^m is a linear combination of the mth element of the basis functions in the set $\{\chi_k\}_{k=1}^{\widehat{K}}$, i.e., $\alpha_{ij}^m = \sum_{k=1}^{\widehat{K}} \sigma_{ijk} \, \chi_k^{(m)}$, where $\chi_k^{(m)} = \langle \chi_k, \mathbf{e_m} \rangle$. This generalizes (5) to the two-dimensional case and avoids to vectorize the data structures $\mathbf{X}_m$.

3.4 Algorithms for tensor decompositions

Clearly, the approximation accuracy of (11) and (13) depend on the choice of basis vectors φ_i, ψ_j, and χ_k, their ordering and the elements in the core tensor. There exist many algorithms to select these bases in such a way that the approximation error (12) is small or minimized. The problem of finding these sets can be formulated as the optimization problem

$$\min_{\{\varphi_i\}_{i \le \widehat{I}}, \{\psi_j\}_{j \le \widehat{J}}, \{\chi_k\}_{k \le \widehat{K}}} \left\| \mathscr{S} - \sum_{i=1}^{\widehat{I}} \sum_{j=1}^{\widehat{J}} \sum_{k=1}^{\widehat{K}} \sigma_{ijk} \varphi_i \otimes \psi_j \otimes \chi_k \right\|_F, \tag{14}$$

which is to be solved subject to the constraint that the basis elements $\{\varphi_i \mid 1 \le i \le \widehat{I}\}$, $\{\psi_j \mid 1 \le j \le \widehat{J}\}$ and $\{\chi_k \mid 1 \le k \le \widehat{K}\}$ are orthonormal.

This problem has an analytic solution only for the case where $(\hat{I}, \hat{J}, \hat{K}) = (1, 1, 1)$. For all other cases one has to resort to numerical approximations. Several algorithms have been proposed to compute tensor decompositions using orthonormal basis functions. The *High Order SVD* (HOSVD) proposed by [11] was the first extension of the classical SVD to the spatial-temporal case and the methodology is based on an unfolding procedure of tensors. The high order orthogonal iteration (HOOI) by [12], the Tensor SVD proposed by [40], maximum singular value modal rank (MSVM), and the single directional modal-rank decomposition (SDM) by [33] compute singular values (elements of the core tensor) and basis vectors in a sequential way, where the singular values and vectors depend on a search direction at every decomposition level $(\widehat{I}, \widehat{J}, \widehat{K})$. The tensor SVD, MSVM, and SDM algorithms keep the tensor structure intact in such a decomposition procedure. In this paper, we consider the *Tucker modal-rank* type of decomposition,

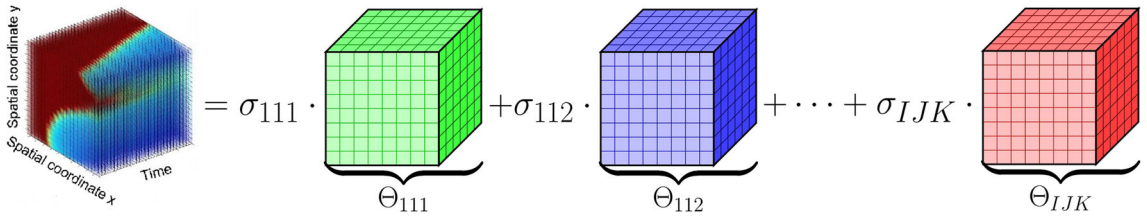

Fig. 2 Schematic description for the truncation of a Tucker decomposition of a 3D tensor

see [22], which achieves orthonormal sets $\{\varphi_i\}_{i=1}^{I}$, $\{\psi_j\}_{j=1}^{J}$ and $\{\chi_k\}_{k=1}^{K}$.

There are several tensor toolboxes available for the Matlab platform, like the Matlab Tensor toolbox, see [5] and the Tensorlab, see [38]. In this work, we have used an HOSVD implementation using tensor operations from [5].

3.5 Tensor approximation of reservoir flow patterns

Let us exemplify the concept of signal approximation and compression of reservoir flow patterns through tensor decompositions. In the framework of multidimensional (tensor) approximations, the sets of orthonormal basis functions $\{\varphi_i\}_{i=1}^{I}$, $\{\psi_j\}_{j=1}^{J}$ represent the most relevant spatial correlations independently for each spatial coordinate. The coordinate independence can be exploited to approximate flow patterns which have a richer variability in a certain coordinate, as it would be the case for flow patterns in channelized reservoirs.

We consider a 2 facies, 2D oil reservoir with a square geometry of length $L = 3000$ m, one layer of 10 m thick. The numerical model of one realization has 3600 grid blocks of size 50 m $\times$ 50 m. A description of the physical parameters, wells configuration, and a link to the data files can be found in [39]. The sequential solvers of MRST, see [24], have been used to solve the pressure and saturation equations and the production has been simulated for a period of 15 years, time step of 5 days. There are four water injectors, and each of them injects at a rate of 600 m^3/day, and the producers operate at 150 bar. We collect $K = 1095$ time steps for the temporal evolution of the oil saturation. Then, we construct a 3D tensor $\mathcal{S}$ of size $60 \times 60 \times 1095$, where the x, y, and temporal dimensions correspond to the first, second, and third tensor coordinate accordingly. Hence, the tensor $\mathcal{S}$ can be decomposed as in Eq. (9):

$$\mathcal{S} = \sum_{i=1}^{60}\sum_{j=1}^{60}\sum_{k=1}^{1095} \sigma_{ijk}\varphi_i \otimes \psi_j \otimes \chi_k. \tag{15}$$

The reservoir flow pattern in tensor $\mathcal{S}$ is described by $I = 60$ basis functions of size 60×1 in the x coordinate, $J = 600$ basis functions of size 60×1 for the y coordinate, and $K = 1095$ basis functions of size 1095×1 for the temporal dimension. We compute low-rank approximations $\widehat{\mathcal{S}}$ of the original flow pattern by truncating the sums in Eq. (15). The purpose of this example is to study the effect of decreasing the number of spatial basis functions for the approximation, and therefore the number of temporal basis functions are fixed to $\{\chi_k\}_{k=1}^{\widehat{K}=5}$. For the first approximation, we select basis $\{\varphi_i\}_{i=1}^{\widehat{I}=20}$ for the x coordinate and basis $\{\psi_j\}_{j=1}^{\widehat{J}=20}$ for the y coordinate, leading to a modal rank approximation of $(20, 20, 5)$. For the second approximation, we select $\{\varphi_i\}_{i=1}^{\widehat{I}=10}$ basis for x and $\{\psi_j\}_{j=1}^{\widehat{J}=10}$ basis for y, leading to a modal rank approximation of $(10, 10, 5)$. Time snapshots of the reservoir simulation and the approximations are depicted in Fig. 3.

From Fig. 3, it is clear that decreasing the number of spatial basis functions would affect the quality of the approximations. This can be quantified by using (12) to derive the relative proximity ν of the approximations $\widehat{\mathcal{S}}$ with respect to the original tensor $\mathcal{S}$: $\nu = \left(1 - \frac{\|\mathcal{S}-\widehat{\mathcal{S}}\|_F}{\|\mathcal{S}\|_F}\right) \times 100$. Here, we consider modal rank approximations of the type $(r, r, 5)$, for $r = 1, \cdots, 60$, and the relative proximity as a function of r is presented in Fig. 4. For the flow patterns depicted in Fig. 3, it is observed that the approximation $(10, 10, 5)$ preserves almost 85% of the features of $\widehat{\mathcal{S}}$, while the approximation $(20, 20, 5)$ preserves more than 90%.

When fixing $\widehat{K} = 5$, only 0.46% of the total number of basis function for the temporal domain is used, while achieving a maximum proximity of 94%. For this example, that fact indicates that the temporal dynamics can be explained with a very small amount of the information contained in $\mathcal{S}$. The amount of information required to construct an approximation can be quantified by summing the size in memory of the constitutive elements in (11): $\{\varphi_i\}_{i=1}^{\widehat{I}}$, $\{\psi_j\}_{j=1}^{\widehat{J}}$, $\{\chi_k\}_{k=1}^{\widehat{K}=5}$, and the corresponding core tensor Σ. Similarly, we consider modal rank approximations

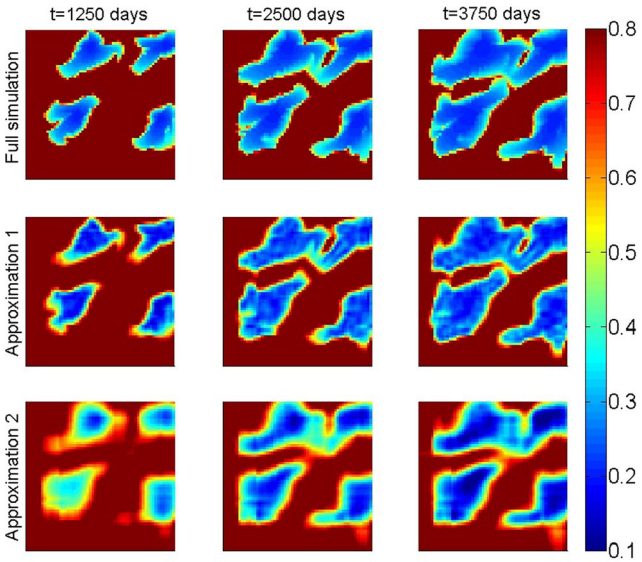

Fig. 3 Oil-water front with tensor approximations. Approximation 1 has modal rank $(20, 20, 5)$. Approximation 2 has modal rank $(10, 10, 5)$. *Colors* represent oil saturation

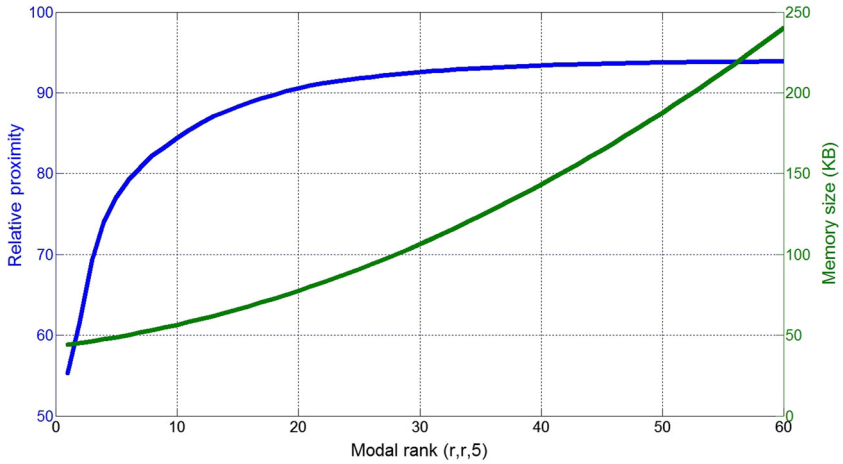

Fig. 4 *Blue-Left axis*: Relative proximity $\nu(r)$. *Green-Right axis*: Size in memory of the modal rank approximation $(r, r, 5)$ as function of r

of the type $(r, r, 5)$, for $r = 1, \cdots, 60$, and the information is presented in Fig. 4. The size in memory of the tensor $\mathscr{S}$ is 30.08 MB, while the approximations $(20, 20, 5)$ and $(10, 10, 5)$ of Fig. 3 have a size of 79.38 and 57.78 KB respectively. These findings indicate that the reservoir flow pattern in $\mathscr{S}$ can be approximated using only 0.25% of its original information, while achieving relative proximities higher than 90%. This experiment suggests that more than 99% of the information contained in $\mathscr{S}$ is redundant for the purpose of flow characterization.

3.6 4D Tensors: an approach for handling multiple realizations

In the multilinear framework, it is possible to define an additional coordinate, where an index that links the dynamical behavior (the reservoir flow pattern) to its corresponding model is assigned to every realization. In this subsection, we restrict our attention to 2D cartesian grids, without losing generality for 3D geometries. For the case where we have an ensemble of R realizations and their corresponding flow patterns, the full data set is described by two spatial coordinates, the temporal coordinate and a coordinate for the realizations, i.e., a 4D tensor $\mathscr{S}$ of size $I \times J \times K \times R$, with I, J the dimension of the spatial coordinates x and y, K the dimension of the temporal coordinate, i.e., the number of time steps, and R the number of reservoir models, which constitutes the size of the ensemble to be characterized. A schematic representation of such a data structure is depicted in Fig. 5. In analogy to the Eq. (9), the 4D tensor $\mathscr{S}$ has a Tucker decomposition of the form:

$$\mathscr{S} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{r=1}^{R} \sigma_{ijkr}\varphi_i \otimes \psi_j \otimes \omega_k \otimes \chi_r, \tag{16}$$

where the orthonormal basis vectors $\{\varphi_i \mid 1 \leq i \leq I\}$, $\{\psi_j \mid 1 \leq j \leq J\}$ span the spatial coordinates, $\{\omega_k \mid 1 \leq k \leq K\}$ spans the temporal space and $\{\chi_k \mid 1 \leq k \leq R\}$ spans the model space. The reservoir flow patterns of the mth realization $\mathscr{X}_m$ is then represented as a *tensor* of dimension $I \times J \times K$ which is approximated similar as in (13), and

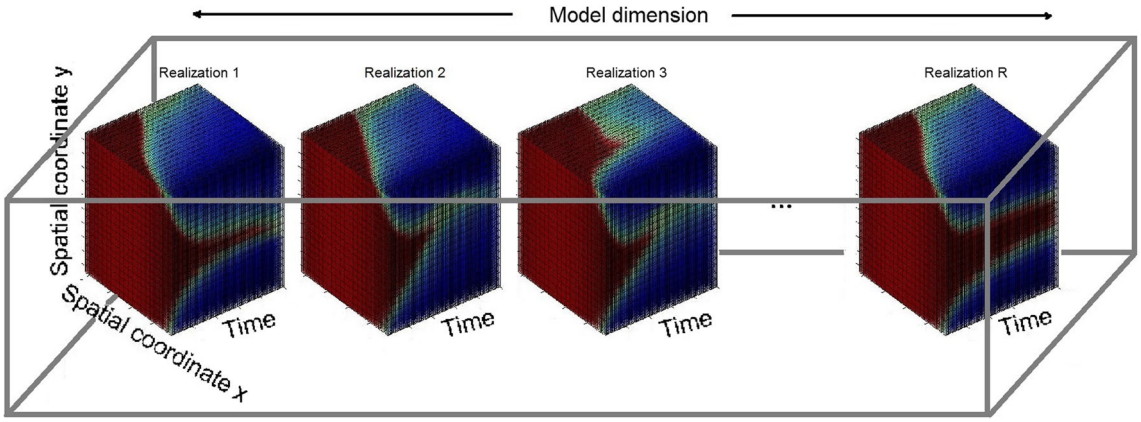

Fig. 5 Schematic interpretation of a 4D tensor of reservoir flow patterns

results in the approximate sample $\widehat{\mathscr{X}_m}$ of the flow patterns defined by the order-3 tensor

$$\begin{aligned}\widehat{\mathscr{X}_m} &= \widehat{\mathscr{S}}(\cdot,\cdot,\cdot,\mathbf{e_m}) \\ &= \sum_{i=1}^{\widehat{I}}\sum_{j=1}^{\widehat{J}}\sum_{k=1}^{\widehat{K}}\sum_{r=1}^{\widehat{R}}\sigma_{ijkr}\langle\chi_r,\mathbf{e_m}\rangle\,\varphi_i\otimes\psi_j\otimes\omega_k \\ &= \sum_{i=1}^{\widehat{I}}\sum_{j=1}^{\widehat{J}}\sum_{k=1}^{\widehat{K}}\alpha_{ijk}^{m}\varphi_i\otimes\psi_j\otimes\omega_k, \end{aligned} \tag{17}$$

where $\mathbf{e_m}$ is the mth standard unit vector in $\mathbb{R}^R$ and where $\alpha_{ijk}^m := \sum_{r=1}^{\widehat{R}}\sigma_{ijkr}\langle\chi_r,\mathbf{e_m}\rangle$ is a real-valued coefficient in the expansion (17).

This expansion shows explicitly the way how the information is distributed in the decomposition. Clearly, the tensor $\varphi_i\otimes\psi_j\otimes\omega_k$ contains the spatial-temporal correlations that are shared by all the set of reservoir flow patterns in the ensemble. What makes a reservoir flow pattern $\mathscr{X}_m$ distinct from others is the selection of the mth element of the basis functions for the model coordinate $\langle\chi_r,\mathbf{e_m}\rangle$ and subsequently the coefficients α_{ijk}^m.

As it was indicated previously for the 3D case, the coefficient α_{ijk}^m is a linear combination of the mth element of all the basis functions in the set $\{\chi_r\}_{r=1}^{\widehat{R}}$, i.e., $\alpha_{ijk}^m = \sum_{r=1}^{\widehat{R}}\sigma_{ijkr}\,\chi_r^{(m)}$, and the information that characterizes the dynamical properties of the realizations are embedded into the elements of core tensor σ_{ijkr} and the set of basis functions for the model coordinate $\{\chi_r\}_{r=1}^{\widehat{R}}$. This analysis creates the foundations for the definition of low-dimensional representations of the flow profiles in the next subsection.

3.7 Flow-based dissimilarity measures in low-dimensional tensor representations

In order to be able to calculate a dissimilarity measure between two models on the basis of low-dimensional representations, we require the tensor representation of the flow patterns for both models to be expanded with the same basis functions, as in Eq. (17). Therefore, we construct the 4D tensor representation described in Section 3.6, and we introduce a metric space by defining a distance function between two reservoir flow patterns $\mathscr{X}_p$ and $\mathscr{X}_q$ of the realizations p, q as:

$$\begin{aligned} d_{pq} &= \|\mathscr{X}_p-\mathscr{X}_q\|_F \\ &= \left\|\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{r=1}^{R}\sigma_{ijkr}\Big[\langle\chi_r,\mathbf{e_p}\rangle-\langle\chi_r,\mathbf{e_q}\rangle\Big]\varphi_i\otimes\psi_j\otimes\omega_k\right\|_F \\ &= \left\|\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{r=1}^{R}\sigma_{ijkr}\langle\chi_r,\mathbf{e_p}-\mathbf{e_q}\rangle\,\varphi_i\otimes\psi_j\otimes\omega_k\right\|_F, \end{aligned} \tag{18}$$

where $\mathbf{e_p}$, $\mathbf{e_q}$ are the pth and qth standard unit vectors in $\mathbb{R}^R$. Let us define the real-valued coefficient $\delta_{ijk} = \sum_{r=1}^{R}\sigma_{ijkr}\langle\chi_r,\mathbf{e_p}-\mathbf{e_q}\rangle$, which is an element of a tensor $\mathscr{D}$ of dimensions $I\times J\times K$. Hence, (18) can be written as

$$\begin{aligned} d_{pq} &= \left\|\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\delta_{ijk}\,\varphi_i\otimes\psi_j\otimes\omega_k\right\|_F \\ &= \|\mathscr{D}\|_F\|\Phi\|_2\|\Psi\|_2\|\Omega\|_2, \end{aligned} \tag{19}$$

where Φ, Ψ and Ω are column matrices composed by the basis functions $\{\varphi_i \mid 1\le i\le I\}$, $\{\psi_j \mid 1\le j\le J\}$ and $\{\omega_k \mid 1\le k\le K\}$. Due to the orthonormality of the columns of Φ, Ψ, and Ω, we obtain:

$$d_{pq} = \|\mathscr{D}\|_F = \sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}|\delta_{ijk}|^2}. \tag{20}$$

The expression in (20) suggests that the dissimilarity between two reservoir realizations can be approximated by truncation, which corresponds to the expression

$$\widehat{d}_{pq} = \sqrt{\sum_{i=1}^{\widehat{I}}\sum_{j=1}^{\widehat{J}}\sum_{k=1}^{\widehat{K}}|\delta_{ijk}|^2}. \tag{21}$$

In addition, the scalar δ_{ijk} can be expressed in terms of the coefficients α_{ijk}^m in (17), $\delta_{ijk} = \alpha_{ijk}^p - \alpha_{ijk}^q$, and therefore the tensor-based approximation of the flow-based distance between the realizations p and q is defined as $\widehat{d}_{pq} = \sqrt{\sum_{i=1}^{\widehat{I}}\sum_{j=1}^{\widehat{J}}\sum_{k=1}^{\widehat{K}}|\alpha_{ijk}^p-\alpha_{ijk}^q|^2}$. If one stores the set of coefficients α_{ijk}^m as elements of a tensor $\mathscr{A}_m$ of dimension $\widehat{I}\times\widehat{J}\times\widehat{K}$, then

$$\widehat{d}_{pq} = \sqrt{\sum_{i=1}^{\widehat{I}}\sum_{j=1}^{\widehat{J}}\sum_{k=1}^{\widehat{K}}|\alpha_{ijk}^p-\alpha_{ijk}^q|^2} = \|\mathscr{A}_p-\mathscr{A}_q\|_F, \tag{22}$$

where $\widehat{d}_{pq}$ is the pqth element of a distance matrix $\widehat{\mathbf{D}}$. The approximation error of computing the dissimilarity measure using the approximations in (17) is bounded (see [11]).

The approximation of the distance in (22) can be computed based on the tensors $\mathscr{A}_p$ and $\mathscr{A}_q$, and they can be seen as compact representations of the pth and qth reservoir flow patterns. The set of tensors $\mathfrak{A} = \{\mathscr{A}_1, \mathscr{A}_2, \cdots, \mathscr{A}_R\}$ is composed by low-dimensional representations of the reservoir flow patterns for an ensemble of R realizations, and they are used for further steps in flow characterization such as distance visualization and model clustering.

3.8 Discussion

With the spatial-temporal methodology, it is possible to identify the tensor coordinates with richer information content, and it introduces flexibility and extra accuracy when representing flow patterns in low-dimensional spaces and calculating dissimilarity measures. Flow-based dissimilarity measures allow the classification of the different types of flow behavior in an ensemble, and the application of the methods described in this section are useful for model clustering, where the computational complexity limits the classification of full reservoir flow patterns. In the next sections, we apply the concept of flow-based dissimilarity measures in low-dimensional spaces to the flow classification of reservoir models using the tensor approach.

4 A tensor-based workflow for model clustering using flow measures

4.1 Introduction

In reservoir engineering, multiple realizations are used to account for the uncertainty of the rock properties of subsurface, and the industrial practice indicates that despite the fact that realizations look different from the geological perspective, some of them may have similar dynamical performance. In flow classification, we aim to find sets of realizations that share a similar dynamical performance with respect to the spatial-temporal evolution of their corresponding reservoir flow patterns. For that, it is required to compute dissimilarity measures between related flow patterns, a method for visualizing these dissimilarities and a clustering technique to group the models with similar dynamical properties. When using a flow-based dissimilarity measure for model clustering, these steps are constrained by the dimensionality of the data set to be analyzed, and the representation of the reservoir flow patterns in low-dimensional spaces are used for the efficient classification of multiple reservoir realizations. The flow-based approach for dissimilarity measures was introduced in Section 3. Here, we provide the theoretical foundation for the workflow developed in this paper. In Section 4.2, we describe a tensor-based clustering algorithm, and in Section 4.3, we describe the method for visualizing distances based on the distance matrix $\widehat{\mathbf{D}}$.

4.2 k-means tensor clustering

When analyzing data sets, analysts aim to extract patterns, object classification, and data ordering. Thereby, k-means clustering finds groups of data which are similar to one another, partitioning a set of objects into clusters. Let us consider the data objects described in Section 3.7, where the tensor data set $\mathfrak{A} = \{\mathscr{A}_1, \mathscr{A}_2, \cdots, \mathscr{A}_R\}$ is composed by the low-dimensional representations of the flow patterns. In this section, we aim for a partition of the data set $\mathfrak{A}$ into a set of K_c clusters $\mathfrak{C} = \{\mathfrak{c}_1, \mathfrak{c}_2, \cdots, \mathfrak{c}_{K_c}\}$ with corresponding centroid μ_k of dimension $\widehat{I} \times \widehat{J} \times \widehat{K}$, the same size of the elements in the set $\mathfrak{A}$, where $\mathfrak{c}_k \subset \mathfrak{A}$ for $k = 1, \cdots K_c$, such that the variance within each cluster is minimized, see [20]. This operation can be formulated as:

$$\arg\min_{\mathfrak{C}} \sum_{k=1}^{K_c} \sum_{i \in \mathfrak{c}_k} \|\mathscr{A}_i - \mu_k\|_F^2, \quad \text{subject to:} \quad \mu_k = \frac{1}{N_k} \sum_{j \in \mathfrak{c}_k} \mathscr{A}_j, \tag{23}$$

where N_k is the number of elements (size) of the cluster $\mathfrak{c}_k$ and $\sum_{j \in \mathfrak{c}_k} \mathscr{A}_j$ indicates the element-wise sum of the tensors $\mathscr{A}_j$ which have been assigned to the cluster $\mathfrak{c}_k$. The k-means algorithm has NP-hard complexity ([3]), which can be relaxed using heuristic algorithms like the Lloyd's algorithm ([25]). The algorithm has two basic steps: (1) The assignment of every tensor object in $\mathscr{A}$ to the closest cluster centroid, and (2) the re-computation of the centroids using the current cluster membership:

- Initialize cluster centroids $\mu_1, \mu_2, \cdots, \mu_{K_c}$ randomly.
- Repeat until convergence:
  1. Label assignment step: Assign each data point to the nearest centroid. For $j = 1, \cdots, R$ and $k = 1, \cdots, K_c$ perform:
  $$l_j = \arg\min_k \|\mathscr{A}_j - \mu_k\|_F^2. \tag{24}$$
  2. Clusters update: Update the set $\mathfrak{C}$.
  3. Centroids update: Compute the average of the cluster elements.
  $$\mu_k = \frac{1}{N_k} \sum_{i \in \mathfrak{c}_k} \mathscr{A}_i. \tag{25}$$

The selection of the K_c is a user choice; however, there are more systematic methods to determine the initial guess for the number of clusters, see, e.g., [21]. When working with large-scale data sets, it is required to account for the scalability and the computational complexity of the algorithms for data analysis. Lloyd's algorithm has linear computational complexity $\mathcal{O}(t \cdot K_c \cdot R \cdot n)$, where t is the number of iterations needed to converge and n is the size of the objects to be clustered respectively. From the complexity point of view, the application of the k-means tensor clustering algorithm is constrained by n, i.e., the dimensionality of the objects to be clustered. Particularly, the fact that

the reservoir flow patterns are large-scale data structures ($n \sim 10^6$) poses a challenge for the state-of-the-art clustering algorithms, and limits the use of k-means for flow-based classification.

For the tensor-based clustering approach, the computational effort will be dominated by the tensor decomposition with computational complexity $\mathcal{O}(\max\{I^3, J^3, K^3, R^3\})$, see, e.g., [11, 15]. A direct clustering on the basis of the full reservoir flow patterns would require a computational effort of order $O(t \cdot K_c \cdot R \cdot I \cdot J \cdot K)$ which will generally be two to three orders of magnitude higher than the former approach, and it is exemplified in the next section.

4.3 Visualization of dissimilarities

For the visualization of dissimilarities between a set of reservoir flow patterns, we determine their coordinates in a metric space using multidimensional scaling (MDS). For a detailed description, we refer to [7]. For the application of MDS in uncertainty quantification, we refer to [32, 35] and [8]. MDS uses the SVD to determine a low-order set of dimensionless directions in which the relative distances between the objects can be efficiently represented. In particular when considering just two or three of the most relevant directions, it is possible to represent the distances between the objects graphically.

4.4 A workflow for model clustering using flow measures

In this subsection, we use the multilinear algebra methods described in this paper to find clusters of models with similar dynamical properties. The developed methodology uses the concept of flow-based dissimilarity measures, computed in low-dimensional spaces to determine the dynamical similarities between reservoir models, by exploiting the tensor structure of the reservoir flow patterns. The purpose of this workflow is to estimate the closeness between two or multiple realizations with respect to a performance indicator relevant to the CLRM framework. The inputs for the workflow are:

- R: The number of realizations.
- $\mathbf{X}_i$: Time snapshots of the reservoir flow patterns ($i = 1, \cdots, R$).
- K_c: The number of clusters.
- A predefined production strategy $u(t)$.

The procedure is described as follows:

1. Reservoir simulation: Simulate the flow patterns for the set of R realizations using $u(t)$.
2. Tensor formulation: Store the reservoir flow patterns of all the realizations in a tensor $\mathscr{S}$ as described in Section 3.2.
3. Decomposition: Compute the tensor decomposition of $\mathscr{S}$ as in Eq. (16), using the algorithms described in Section 3.4.
4. Low-dimensional characterization: Construct the low-dimensional representation of the flow profiles $\mathfrak{A} = \{\mathscr{A}_1, \mathscr{A}_2, \cdots, \mathscr{A}_R\}$, as described in Section 3.7.
5. Dissimilarity: Compute the distances described in Eq. (22).
6. Clustering: Group the data set $\mathfrak{A}$ into clusters as described in Section 4.2.
7. Visualization: Construct an MDS map to visualize clusters as described in Section 4.3.

The output of the workflow is the set of K_c clusters $\mathfrak{C} = \{\mathfrak{c}_1, \mathfrak{c}_2, \cdots, \mathfrak{c}_{K_c}\}$, which groups the types of dynamical responses of R reservoir realizations. This classification can be further used to create flow-relevant ensembles, where few reservoir models are selected to capture the most relevant dynamical responses of the original set of realizations.

It has to be noted that, if we consider the saturation-based dissimilarity measure, then the presented model clustering method requires a full simulation of all model realizations in the ensemble. In a general CLRM workflow, including robust optimization and/or value of information assessment, this burden will generally be outweighted by the advantages of performing optimization over a considerably reduced ensemble. In the current procedure, the clustering can be done once and "off-line." Nevertheless, if a full simulation of all reservoir realizations is considered unfeasible, the proposed method and techniques in this paper can still be applied e.g. on the basis of time-of-flight maps rather than saturation maps (see, e.g., [40]), thereby considerably reducing the simulation efforts.

5 Application case

In this section, the workflow for model clustering presented in Section 4 is applied to a set of channelized reservoirs, and we analyze the performance of the spatial-temporal approach using flow-based dissimilarity measures. Channelized reservoirs present a challenge for field development plans, because moderate changes in well configurations may lead to very high variations in the resulting reservoir flow patterns. Let us consider an ensemble of $R = 100$, 3D reservoir models with a geological structure consisting on a network of fossilized meandering channels of high permeability. The data set has been uploaded to the 4TU.Datacentrum repository and can be accessed by external users, see [19] for the physical parameters of the models. The reservoir size is 480 m $\times$ 480 m $\times$ 28 m with 7 geological layers, and it is composed by 25,200 grid blocks 8 m $\times$ 8 m $\times$ 4 m in size.

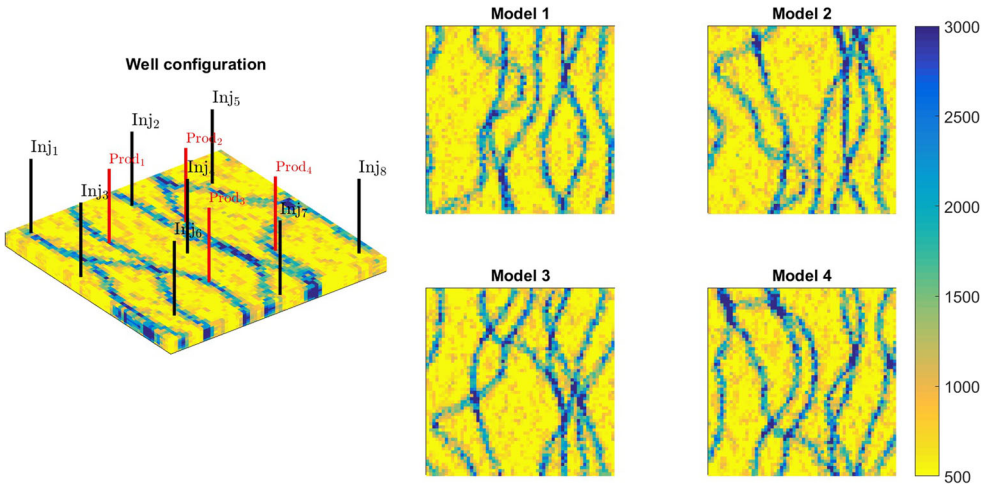

Fig. 6 Well configuration and samples of permeability fields from the ensemble. Color scale in mDarcy

We have used a rectangular-shaped geometry instead of the egg-shaped reservoir described originally in [19]. The well configuration is composed of eight injectors and four producers. A view of some geological realizations is depicted in Fig. 6.

5.1 Generation of the reservoir flow patterns

We have simulated the reservoir flow patterns of the $R = 100$ realizations, which correspond to the spatial-temporal evolution of the oil saturation. The sequential solvers of MRST, see [24], have been used to solve the pressure and saturation equations, and the production has been simulated for a period of 10 years with a time step of 30 days, i.e., $K = 122$ time steps. The water injection rates are fixed at 79.5 m^3/day for all the injectors and the bottom-hole pressures are fixed at 395 bar for all the producers.

5.2 Low-dimensional tensor representation of the reservoir flow patterns

The data structure that contains the reservoir flow patterns for the ensemble can be stored in a 5D tensor of size $I \times J \times Z \times K \times R$ with $I = 60$ the dimension of the x coordinate, $J = 60$ the dimension of the y coordinate, $Z = 7$ layers, $K = 122$ time steps and $R = 100$ realizations. The tensor $\mathscr{S}$ is decomposed similarly to the decomposition in (16), while augmenting a coordinate for geological layers. Hence, the reservoir flow patterns corresponding to the mth realization can be described as:

$$\widehat{\mathscr{X}_m} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{z=1}^{Z}\sum_{k=1}^{K} \alpha_{ijzk}^{m}\varphi_i \otimes \psi_j \otimes \nu_z \otimes \omega_k, \tag{26}$$

where $\{\nu_z\}_{z=1}^{Z}$ are the set of basis functions for the layers coordinate and $\alpha_{ijzk}^{m} := \sum_{r=1}^{R} \sigma_{ijzkr}\langle \chi_r, \mathbf{e_m}\rangle$. For the low-dimensional approximation of $\mathscr{S}$, we truncate the number of the basis functions in every coordinate such that the approximation error described in (12) is relatively small. We have set the number of basis functions for the layers coordinate to be $\hat{Z} = 2$, and for the temporal coordinate to be $\hat{K} = 2$. From the expression in (17), it is inferred that the number of basis functions for the model coordinate does not affect the number of parameters α_{ijzk} required to describe a flow pattern, and thus we select $\hat{R} = 100$ basis functions for the model coordinate. In order to choose an adequate number of spatial basis functions for the x and y coordinates,

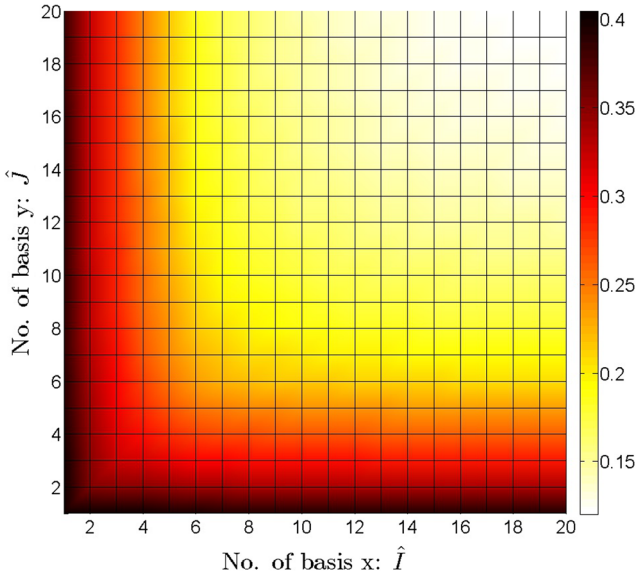

Fig. 7 Approximation error $e_r(\hat{I}, \hat{J}) = \frac{\|\mathscr{S} - \widehat{\mathscr{S}}(\hat{I},\hat{J})\|_F}{\|\mathscr{S}\|_F}$

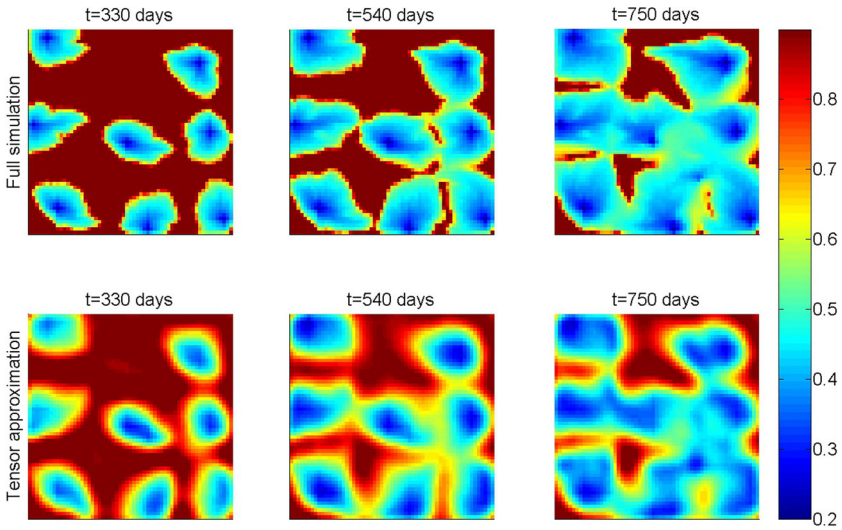

Fig. 8 Snapshots of oil saturation for model 57 (layer 3) with $\hat{I} = 10$, $\hat{J} = 10$, $\hat{Z} = 2$, $\hat{K} = 2$, and $\hat{R} = 100$. *Top:* Reservoir simulation. *Bottom:* Tensor approximation

we perform an error analysis using the approximation error defined in (12).

In Fig. 8, the approximation error as a function of the number of basis functions used for the approximation $\widehat{\mathscr{S}}$ is presented. Using Fig. 8, we can perform a trade-off between accuracy and the amount of information required for the approximation. The truncation criterion is defined by the user, and we accept tensor approximations with relative errors lower than 20% with respect to the original tensor $\mathscr{S}$. By choosing a truncation $\hat{I} = 10$ and $\hat{J} = 10$ we achieve a relative error of 17.6% with respect to the flow patterns from the full simulation. Figure 8 displays an almost symmetric shape, and shows that the approximation error tends to zero as we increase the number of basis functions in both coordinates.

Snapshots of the approximation for one of the realizations are depicted in Fig. 7. Despite that the oil/water front is not as sharp as in the reservoir simulation, the tensor approximation is able to capture the relevant flow patterns as time evolves. These approximations allow the characterization of the reservoir flow patterns in low-dimensional spaces. As was described in the previous section, the reservoir flow pattern of the mth realization $\mathscr{X}_m$ (of size $I \times J \times Z \times K$, i.e. 3.074×10^6 grid-block oil saturations) is characterized in a low-dimensional space by the tensor $\mathscr{A}_m$ of size $\widehat{I} \times \widehat{J} \times \widehat{Z} \times \widehat{K}$ composed by the set of coefficients α_{ijzk}, i.e., 400 coefficients. This low-dimensional characterization represents a reduction of 99.9% of the amount of information necessary for further classification analysis.

5.3 Model clustering and visualization

The low-dimensional tensor representation in (26) and the methods described in Section 4.4 are used for model clustering. In the previous section, we have derived a set of tensors $\mathfrak{A} = \{\mathscr{A}_1, \mathscr{A}_2, \cdots, \mathscr{A}_R\}$ which are low-dimensional representations of the reservoir flow patterns for an ensemble of R realizations. Here, we use them for constructing model clusters with similar reservoir flow patterns.

The k-means clustering algorithm described in Section 4.2 has been used to classify the set $\mathfrak{A}$. The algorithm is provided with the number of clusters $K_c = 7$,

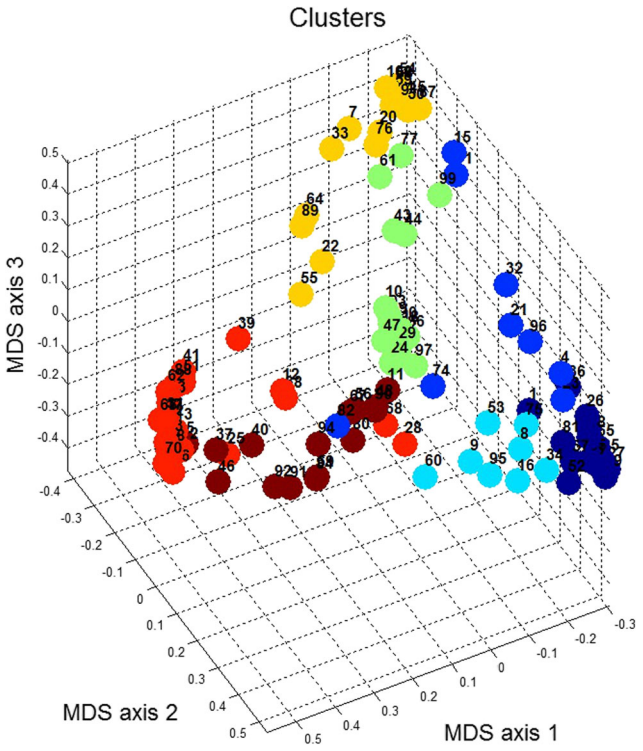

Fig. 9 MDS plot. *Color* represent flow-based clusters. *Numbers* are assigned to all the realizations

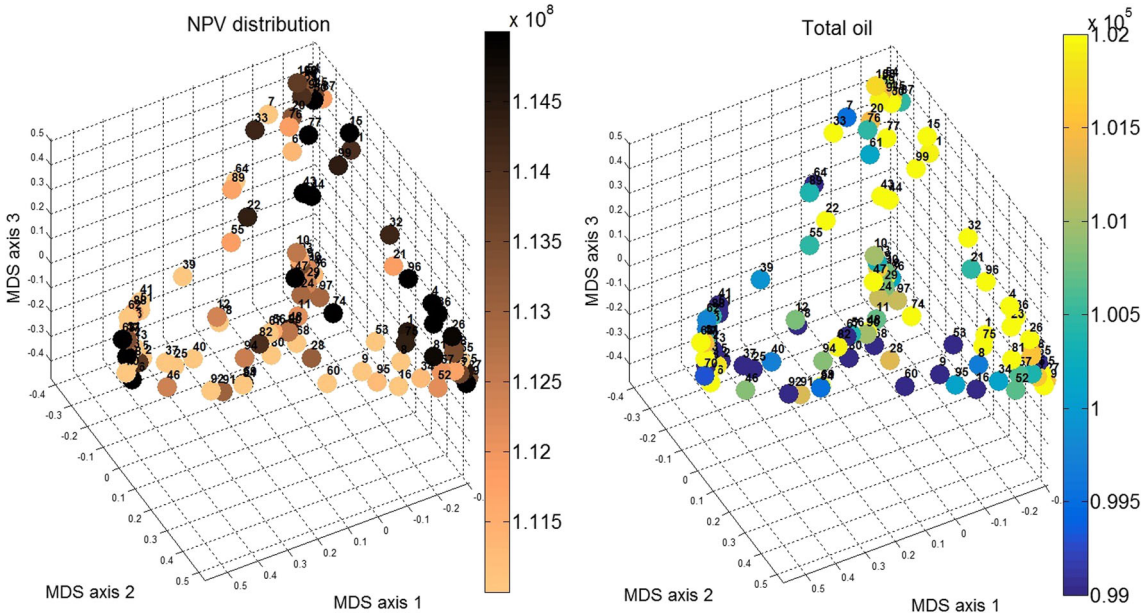

Fig. 10 MDS plot. *Left*: *Color* represents NPV (USD). *Right*: *Color* represents total oil production (stb). *Numbers* are assigned to all the realizations

which have been selected by visual inspection of the MDS plots depicted in Fig. 9. A predefined initial condition for all the centroids $\mu_0 = 0.1 \cdot \mathscr{E}$, where $\mathscr{E}$ is a tensor of size $10 \times 10 \times 2 \times 2$ with all the elements equal to 1.

To visualize the clusters, the MDS map is constructed using a distance matrix $\widehat{\mathbf{D}}$ based on the approximated dissimilarity function defined in (22). In the MDS plot, every dot corresponds to a low-dimensional representation of the reservoir flow patterns for one realization, and the colors indicate clusters. The MDS map in Fig. 9 is a 3D projection of a higher-dimensional space, and the first three axes represent 72% of the total variability. The clusters depicted in the MDS plot are visually separated and the plot has a tetrahedron shape. Big clusters are found in the corners of the tetrahedron, indicating four predominant and different types of flow patterns.

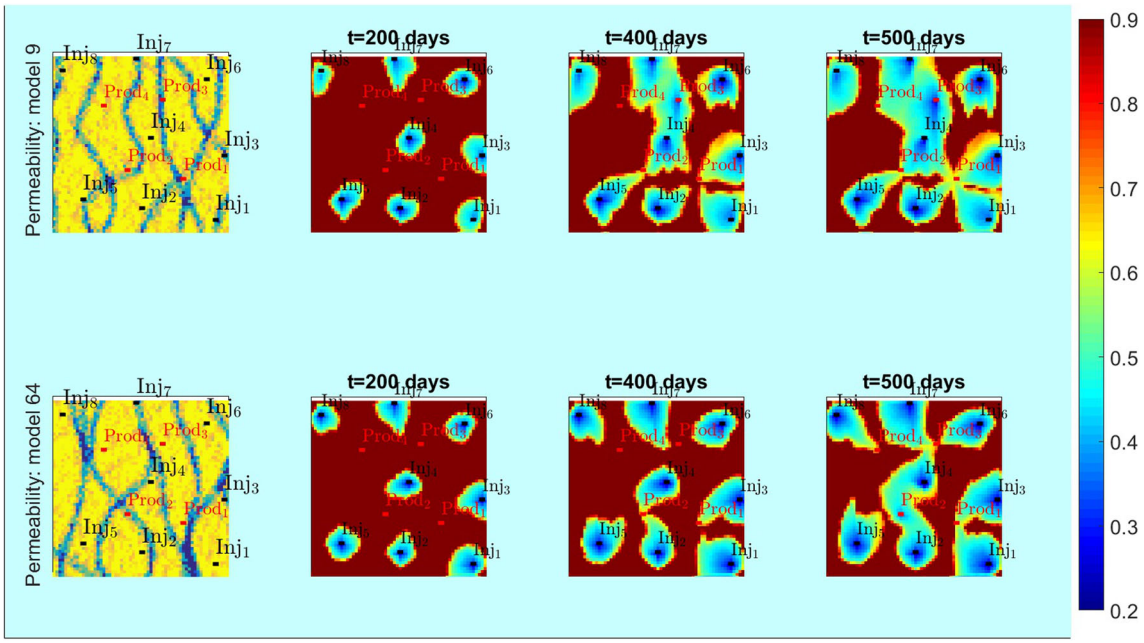

Fig. 11 Snapshots of oil saturation (*top layer*) of sample models from cluster 4

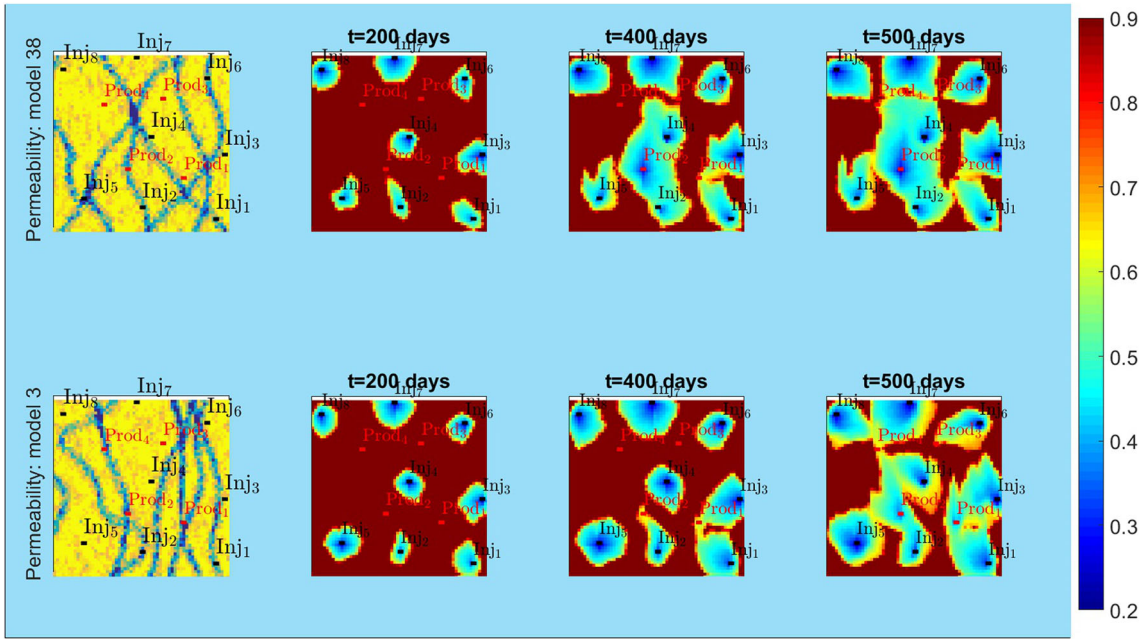

Fig. 12 Snapshots of oil saturation (*top layer*) of sample models from cluster 2

The distribution of NPV and total oil for the clusters can be visualized in the MDS plots of Fig. 10. From visual inspection, patterns of similar NPV, and total oil production can be detected for models which belong to the same cluster and are close in the flow patterns, and there is a positive correlation of the model distances with the target variables.

Typical flow patterns of these clusters are depicted in Figs. 11, 12, 13, and 14.

To describe the observations, let us analyze the type of reservoir flow patterns classified in some of the clusters. In Fig. 11, the flow patterns corresponding to the realizations 9 and 64 indicate a large connectivity between the injectors 2, 4, and 7 with the producers 2 and 3 resulting in flow patterns with an elongated shape in the y coordinate. In Fig. 12, the flow patterns of the realizations 3 and 38 present a large connectivity between injectors 1, 2, 3, and 4 with

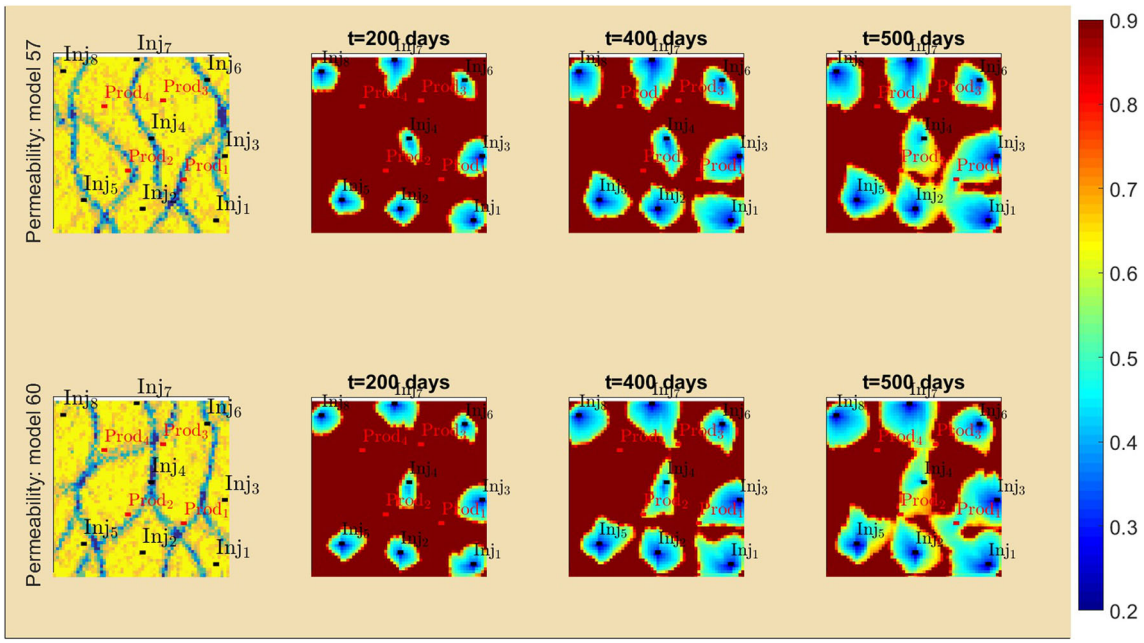

Fig. 13 Snapshots of oil saturation (*top layer*) of sample models from cluster 1

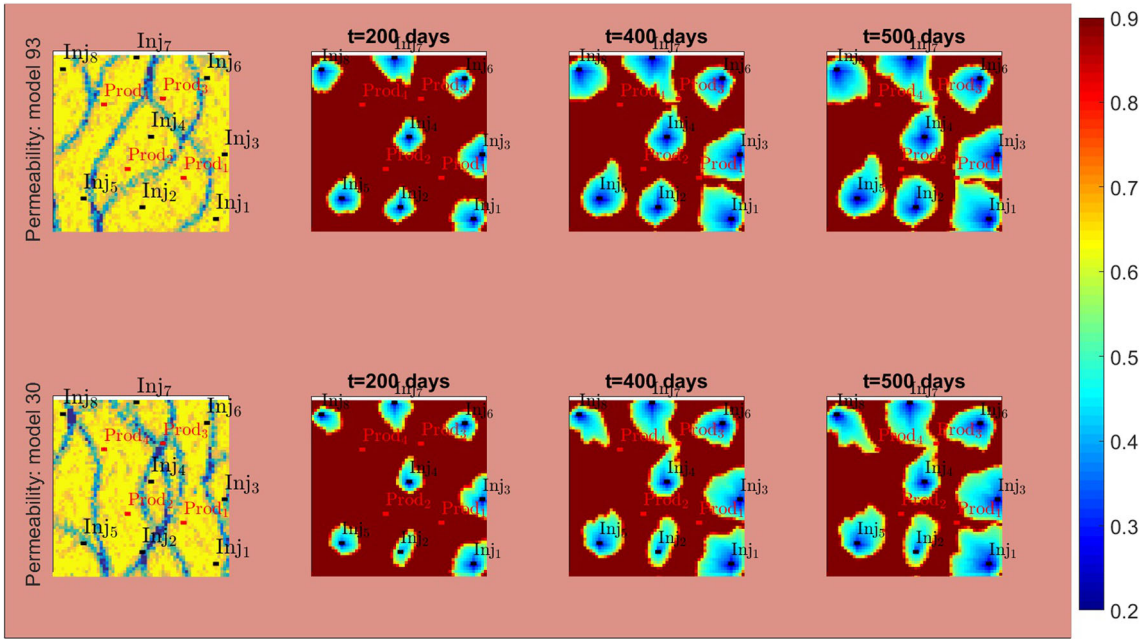

Fig. 14 Snapshots of oil saturation (*top layer*) of sample models from cluster 3

producers 1, 2, and 3 resulting in a rounded flow pattern in the center of the reservoir. In Fig. 13, the flow patterns indicate a large connectivity between injectors 1, 2, and 3 with producer 1, while the flow patterns corresponding to the cluster in Fig. 14 do not exhibit such connectivity. The results depicted in Figs. 11, 12, 13, and 14 confirm the effectivity of the spatial-temporal workflow for model clustering using flow-based dissimilarity measures.

In this application case, the computational complexity of the tensor decomposition comes down to $\mathcal{O}(K^3) = \mathcal{O}(122^3) \sim \mathcal{O}(\mathbf{2 \cdot 10^6})$. Since the flow patterns have a reduced-order representation composed of $n = 400$ elements, the application of Lloyd's k-means clustering algorithm has a computational complexity of order $\sim \mathcal{O}(\mathbf{0.0004 \cdot 10^6})$. Alternatively, the brute force calculation of all pairwise distances between the original feature vectors, requires a number of computations of the order $\mathcal{O}(\frac{1}{2}R^2 \cdot I \cdot J \cdot Z \cdot K)$, which for the example case comes down to $\mathcal{O}(\frac{1}{2}100^2 \cdot 60 \cdot 60 \cdot 7 \cdot 122) \sim \mathcal{O}(\mathbf{1.5 \cdot 10^{10}})$, without considering the clustering step. Applying Lloyd's clustering algorithm to the original features has a computational complexity of order $\mathcal{O}(it \cdot K_c \cdot \frac{1}{2}R^2 \cdot I \cdot J \cdot Z \cdot K) = \mathcal{O}(it \cdot 7 \cdot 5000 \cdot 60 \cdot 60 \cdot 7 \cdot 122) \sim \mathcal{O}(it \cdot \mathbf{10^{11}})$. From these figures, it is clear that the computational advantages of the tensor reduction step are substantial. This computational gain will greatly facilitate the implementation of clustering algorithms for problems of practical size.

5.4 NPV and oil production in the clusters

Previously, we have discussed the fact that models with similar outputs such as NPV or production rates might have very different reservoir flow patterns. In the context of model clustering with flow-based dissimilarity measures, it is expected that the models within a cluster share similar flow patterns. We anticipate that the NPVs and total oil productions are similar as well, based on the fact that models with similar flow patterns might have similar

Table 1 Average and standard deviation of NPV and total oil for each of the seven clusters and for the full ensemble

Cluster		NPV (10^6 USD)		Total Oil (10^3 stb)	
No.	N_k: Cluster size	μ_{npv}	σ_{npv}	μ_{oil}	σ_{oil}
1	14	113.64	2.33	101.67	1.55
2	9	115.44	2.22	102.86	1.48
3	8	110.11	2.24	99.32	1.49
4	15	114.02	2.47	101.92	1.64
5	17	113.13	2.20	101.33	1.46
6	22	111.47	2.54	100.22	1.69
7	15	111.03	2.54	99.93	1.69
Ensemble	100	112.62	2.80	100.99	1.86

Table 2 Production strategies

Strategy	Injection rates	Deviation
Base case	$r = 79.5\ \mathrm{m}^3/\mathrm{day}$	–
2	$r_{\mathrm{odd}} = r - \Delta_1, r_{\mathrm{even}} = r + \Delta_1$	$\Delta_1 = 4\ \mathrm{m}^3/\mathrm{day}$
3	$r_{\mathrm{odd}} = r + \Delta_1, r_{\mathrm{even}} = r - \Delta_1$	$\Delta_1 = 4\ \mathrm{m}^3/\mathrm{day}$
4	$r_{\mathrm{odd}} = r - \Delta_2, r_{\mathrm{even}} = r + \Delta_2$	$\Delta_2 = 10\ \mathrm{m}^3/\mathrm{day}$
5	$r_{\mathrm{odd}} = r + \Delta_2, r_{\mathrm{even}} = r - \Delta_2$	$\Delta_2 = 10\ \mathrm{m}^3/\mathrm{day}$

evolution of the oil saturation and pressure patterns, generating close water breakthrough times and rates at the production wells. For the application case considered in this section, we compute the undiscounted NPV as in [18], with $r_o = 55$ USD/stb the oil price, $r_{wi} = r_{wp} = 2$ USD/stb the cost associated to water injection and production. The range of NPV for the ensemble is $\omega_{NPV} := [104.82 \times 10^6,\ 119.27 \times 10^6]$ USD, with a mean value of $\mu_{NPV} = 112.62 \times 10^6$ USD and standard deviation $\sigma_{NPV} = 2.80 \times 10^6$ USD. The statistical properties (mean and standard deviation) of NPV and total oil production of the flow-based clusters are presented in Table 1. From the table, we conclude that for all seven clusters, the intra-cluster standard deviations of NPV (σ_{npv}) and total oil (σ_{oil}) are smaller than the standard deviations of these variables over the full ensemble. This is an evidence for the statement that the clustering has been done in a way that is relevant with respect to these two performance measures.

5.5 Input dependency of the reservoir flow patterns and dissimilarity measures

One of the possible limitations of the workflow is the nonlinear dependency of the reservoir flow patterns on the type of production strategy. In this section, we compare the results of clustering using flow-based dissimilarity measures for different production strategies. The strategies that will be considered in this section consist on a fixed bottom-hole pressure of 395bar at the producers, while the injection rates are perturbed by ± 5 and $\pm 12.5\%$, as specified in Table 2.

In general, it would be beneficial to have a small sensitivity of the workflow to variations in the control input. In order to assess this sensitivity, we have applied the workflow for flow characterization with the control strategies described in Table 2, and have generated $K_c = 7$ clusters for each strategy. Here, we define the closeness of the clusters with deviated control to the clusters of the base case, as the number of shared models among the clusters. This cluster similarity can be quantified as the percentage of the elements shared with the cluster in the base case. Let $\mathfrak{c}_r^{\mathrm{base}}$ and $\mathfrak{c}_r^n$ be the rth clusters for the base and the nth production strategy. We define the percentage of similarity of the cluster $\mathfrak{c}_r^n$ with respect to the base case $\mathfrak{c}_r^{\mathrm{base}}$ as:

$$p_r = \frac{\mathrm{card}(\mathfrak{c}_r^{\mathrm{base}} \cap \mathfrak{c}_r^n)}{\mathrm{card}(\mathfrak{c}_r^{\mathrm{base}})} \cdot 100, \tag{27}$$

where $\mathrm{card}(\cdot)$ denotes the set cardinality, i.e., the number of elements of a set. In Fig. 15, this percentage of similarity is presented.

The results in Fig. 15 indicate that for small deviations (strategies 2,3), there is a good agreement of the clusters for deviated controls with the classification obtained by the base case, as most of the clusters match the base case with at least 68%, with the exception of cluster 7 for strategy 2. This is expected, as there are no large variations in the reservoir flow patterns for small variations in the controls. For large deviations in the control inputs (strategies 4, 5),

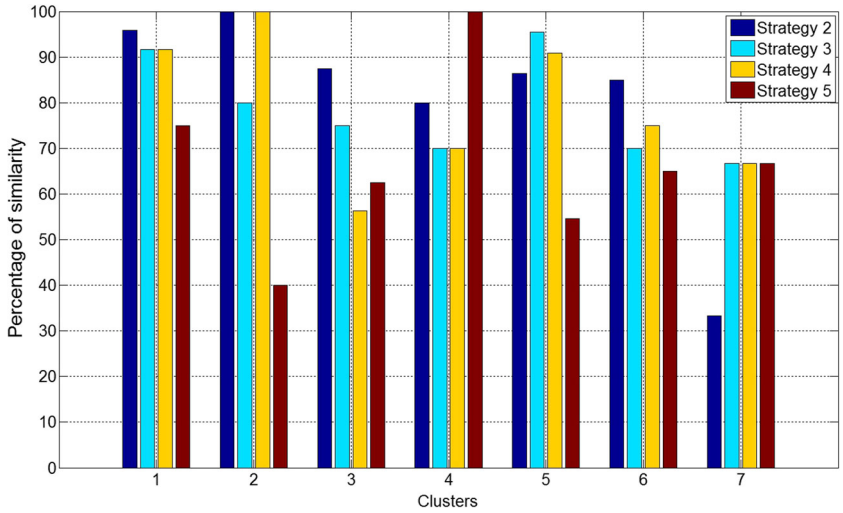

Fig. 15 Model clusters similarity with respect to clusters of the base case

the similarity of the clusters with respect to the base cases decreases, however, there it is still a good match, and the generated clusters have a significant similarity with respect to the classification found in the base case.

The results indicate that the nonlinear dependency of the flow patterns on the control input is an inherent limitation of the workflow. However, this methodology is valid for small deviations around the production strategy.

6 Conclusions

Some relevant advantages have been identified for the proposed methodology of model classification using flow-based dissimilarity measures and tensor representations: Firstly, the spatial structure of the reservoir is preserved, which allows the extraction of spatial correlations from the reservoir flow patterns. Secondly, the spatial correlations are not averaged in time, which is particularly useful for the flow characterization of nonlinear reservoir systems, where the spatial correlations of the reservoir variables are time-variant. Thirdly, a tensor-based representation provides the user with enough flexibility for handling multidimensional reservoir flow patterns and for performing a directional approximation of the data, i.e., keeping the directions where the dynamics have a higher variability. As a consequence, the tensor approximations can represent patterns in the full simulation using only 0.1% of the original information.

Finally, a low-dimensional representation of the reservoir flow patterns allows the implementation of dissimilarity and clustering techniques for reservoir models. The presented clustering technique can be used to construct reduced-sized ensembles for instance for applying robust life cycle optimization [37], value of information assessment [6], or well placement optimization while the original uncertainty in the ensemble of reservoir models is captured by a reduced set of ensemble members, chosen after flow-relevant clustering of the realizations, and thereby leading to substantial computational benefits.

Acknowledgements We acknowledge the discussions with Dr. Tzu-hao Yeh from the Quantitative Reservoir Management group at Shell for his views on the potential application of the techniques presented in this paper on field cases. The authors acknowledge financial support from the Recovery Factory program sponsored by Shell Global Solutions International.

References

1. Afra, S., Gildin, E.: Permeability parametrization using higher order singular value decomposition (hosvd). In: 2013 12th International Conference on Machine Learning and Applications (ICMLA), vol. 2, pp. 188–193. IEEE (2013). doi:10.1109/icmla.2013.121
2. Afra, S., Gildin, E., Tarrahi, M.: Heterogeneous reservoir characterization ussing efficient parameterization through higher order svd (hosvd). In: American Control Conference, pp. 147–152, Portland, Oregon (2014). doi:10.1109/acc.2014.6859,246
3. Aloise, D., Deshpande, A., Hansen, P., Popat, P.: Np-hardness of euclidean sum-of-squares clustering. Mach. Learn. **75**(2), 245–248 (2009). doi:10.1007/s10,994–009–5103–0
4. Aziz, K., Settari, A.: Petroleum reservoir simulation, vol. 476 Applied Science Publishers London (1979)
5. Bader, B.W., Kolda, T.G., et al.: Matlab tensor toolbox version 2.6. Available online http://www.sandia.gov/tgkolda/TensorToolbox/ (2015)
6. Barros, E.G.D., Van den Hof, P.M.J., Jansen, J.D.: Value of information in closed-loop reservoir management. Comput. Geosci. **20**(3), 737–749 (2016). doi:10.1007/s10,596–015–9509–4
7. Borg, I., Groenen, P.J.F.: Modern multidimensional scaling: Theory and applications. Springer. doi:10.4324/9780203767719 (2005)
8. Caers, J., Park, K., Scheidt, C.: Modeling uncertainty of complex earth systems in metric space. In: Handbook of Geomathematics, pp. 865–889. Springer (2010). doi:10.1007/978–3–642–01,546–5–29
9. Cardoso, M.A., Durlofsky, L.J., Sarma, P.: Development and application of reduced-order modeling procedures for subsurface flow simulation. Int. J. Numer. Methods Eng. **77**(9), 1322–1350 (2009). doi:10.1002/nme.2453
10. Chen, Y., Oliver, D.S., Zhang, D., et al: Efficient ensemble-based closed-loop production optimization. SPE J. **14**(04), 634–645 (2009). doi:10.2118/112,873–pa
11. De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. **21**(4), 1253–1278 (2000). doi:10.1137/s0895479896305,696
12. De Lathauwer, L., De Moor, B., Vandewalle, J.: On the best rank-1 and rank-$(r_1, r_2, \cdots, r_n)$ approximation of higher-order tensors. SIAM J. Matrix Anal. Appl. **21**(4), 1324–1342 (2000). doi:10.1137/s0895479898346,995
13. Durlofsky, L.J.: Upscaling and gridding of fine scale geological models for flow simulation. In: 8th International Forum on Reservoir Simulation Iles Borromees, Stresa, Italy, pp. 20–24 (2005). https://pangea.stanford.edu/ERE/research/suprihw/durlofsky/upsc_grid_review_ifrs_2005.pdf
14. Gildin, E., Afra, S.: Efficient inference of reservoir parameter distribution utilizing higher order svd reparameterization. In: ECMOR XIV-14th European conference on the mathematics of oil recovery. Catania, Italy (2014). doi:10.3997/2214-4609.20141826
15. Golub, G.H., Van Loan, C.F.: Matrix computations, vol. 3. JHU Press. doi:10.1137/1028073 (2012)
16. Insuasty, E., Van den Hof, P.M.J., Weiland, S., Jansen, J.D.: Tensor-based reduced order modeling in reservoir engineering: An application to production optimization. IFAC-PapersOnLine **48**(6), 254–259 (2015). doi:10.1016/j.ifacol.2015.08.040
17. Jansen, J.D.: A systems description of flow through porous media. Springer Briefs in Earth Sciences, Springer. doi:10.1007/978-3-319-00260-6 (2013)
18. Jansen, J.D., Bosgra, O.H., Van den Hof, P.M.J.: Model-based control of multiphase flow in subsurface oil reservoirs. J. Process Control **18**(9), 846–855 (2008). doi:10.1016/j.jprocont.2008.06.011

19. Jansen, J.D., Fonseca, R.M., Kahrobaei, S., Siraj, M.M., Van Essen, G.M., Van den Hof, P.M.J.: The egg model–a geological ensemble for reservoir simulation. Geosci. Data J. **1**(2), 192–195 (2014). doi:10.1002/gdj3.21
20. Jegelka, S., Sra, S., Banerjee, A.: Approximation algorithms for tensor clustering. In: Algorithmic learning theory, pp. 368–383. Springer (2009). doi:10.1007/978–3–642–04,414–4–30
21. Ketchen, D.J., Shook, C.L.: The application of cluster analysis in strategic management research: an analysis and critique. Strateg. Manag. J. **17**(6), 441–458 (1996). doi:10.1002/(SICI)1097–0266(199,606)17:6<441::AID–SMJ819>3.0.CO;2–G
22. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009). doi:10.1137/07070,111x
23. Krogstad, S.: A sparse basis pod for model reduction of multiphase compressible flow. In: SPE Reservoir Simulation Symposium. The Woodlands, Texas. doi:10.2118/141973-ms (2011)
24. Lie, K.A., Krogstad, S., Ligaarden, I.S., Natvig, J.R., Nilsen, H.M., Skaflestad, B.: Open-source matlab implementation of consistent discretisations on complex grids. Comput. Geosci. **16**(2), 297–322 (2012). doi:10.1007/s10,596–011–9244–4
25. Lloyd, S.: Least squares quantization in pcm. IEEE Trans. Inf. Theory **28**(2), 129–137 (1982). doi:10.1109/TIT.1982.1056,489
26. Markovinovic, R., Jansen, J.D.: Accelerating iterative solution methods using reduced-order models as solution predictors. Int. J. Numer. Methods Eng. **68**(5), 525–541 (2006). doi:10.1002/nme.1721
27. Park, K., Caers, J.: History matching in low-dimensional connectivity-vector space. In: EAGE Petroleum Geostatistics. Cascais, Portugal. doi:10.3997/2214-4609.201403075 (2007)
28. Sarma, P., Chen, W., Xie, J.: Selecting representative models from a large set of models. In: SPE Reservoir Simulation Symposium. The Woodlands, Texas. doi:10.2118/163671-MS (2013)
29. Sarma, P., Durlofsky, L.J., Aziz, K.: Computational techniques for closed–loop reservoir modeling with application to a realistic reservoir. Pet. Sci. Technol. **26**(10–11), 1120–1140 (2008). doi:10.1080/10916460701829,580
30. Scheidt, C., Caers, J.: Representing spatial uncertainty using distances and kernels. Math. Geosci. **41**(4), 397–419 (2009). doi:10.1007/s11,004–008–9186–0
31. Scheidt, C., Caers, J., Chen, Y., Durlofsky, L.: A multi-resolution workflow to generate high-resolution models constrained to dynamic data. Comput. Geosci. **15**(3), 545–563 (2011). doi:10.1007/s10,596–011–9223–9
32. Scheidt, C., Caers, J., et al: Uncertainty quantification in reservoir performance using distances and kernel methods–application to a west africa deepwater turbidite reservoir. SPE J. **14**(04), 680–692 (2009). doi:10.2118/118,740–PA
33. Shekhawat, H.S., Weiland, S.: On the problem of low rank approximation of tensors. In: 21^{st} International Symposium on Mathematical Theory of Networks and Systems. Groningen, The Netherlands. http://fwn06.housing.rug.nl/mtns2014-papers/fullPapers/0386.pdf (2014)
34. Suzuki, S., Caers, J.: A distance based prior model parameterization for constraining solution of spatial inverse problems. Math. Geosci. **40**(4), 445–469 (2008). doi:10.1007/s11,004–008–9154–8
35. Suzuki, S., Caumon, G., Caers, J.: Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization. Comput. Geosci. **12**(1), 105–119 (2008). doi:10.1007/s10,596–007–9063–9
36. Van Doren, J.F.M., Van den Hof, P.M.J., Bosgra, O.H., Jansen, J.D.: Controllability and observability in two-phase porous media flow. Comput. Geosci. **17**(5), 773–788 (2013). doi:10.1007/s10,596–013–55–1
37. Van Essen, G.M., Zandvliet, M.J., Van den Hof, P.M.J., Bosgra, O.H., Jansen, J.D.: Robust waterflooding optimization of multiple geological scenarios. SPE J. **14**(01), 202–210 (2009). doi:10.2118/102,913–ms
38. Vervliet, N., Debals, O., Sorber, L., Barel, M.V., Lathauwer, L.D.: Tensorlab v3.0. Available online http://www.tensorlab.net (2016)
39. Vo, H.X., Durlofsky, L.J.: Data assimilation and uncertainty assessment for complex geological models using a new pca-based parameterization. Comput. Geosci. **19**(4), 747–767 (2015). doi:10.1007/s10,596–015–9483–x
40. Weiland, S., Van Belzen, F.: Singular value decompositions and low rank approximation of tensors. IEEE Trans. Signal Process. **58**(3), 1171–1182 (2010). doi:10.1109/tsp.2009.2034,308
41. Yeh, T.H., Jimenez, E., Van Essen, G., Chen, C., Jin, L., Girardi, A., Gelderblom, P., Horesh, L., Conn, A.R., et al: Reservoir uncertainty quantification using probabilistic history matching workflow. In: SPE Annual Technical Conference and Exhibition. Amsterdam, The Netherlands. doi:10.2118/170893-ms (2014)